

# Enhancing Liver Segmentation: A Deep Learning Approach with EAS Feature Extraction and Multi-Scale Fusion

\*Weimin Wang<sup>1</sup>, Yufeng Li<sup>2</sup>, Xu Yan<sup>3</sup>, Mingxuan Xiao<sup>4</sup>, and Min Gao<sup>5</sup>

<sup>1</sup> Hong Kong University of Science and Technology, HONG KONG, China

<sup>2</sup> University of Southampton, United Kingdom

<sup>3,5</sup> Trine University, United State of America

<sup>4</sup> SouthWest JiaoTong University, China

\*Correspondence should be addressed to Weimin WANG; [wangwaynem1n@gmail.com](mailto:wangwaynem1n@gmail.com)

Received 2 January 2024;

Revised 16 January 2024;

Accepted 30 January 2024

Copyright © 2024 Made Weimin WANG et al. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**ABSTRACT-** Deep learning technology have been broadly used in segmentation tasks of liver. To address the limitation of suboptimal segmentation for small targets, an end-to-end EAS(ECA-Attention and Separable convolution) U-Net is proposed based on deep learning. The basic module employs depthwise separable convolutional modules instead of convolutional modules to reduce the parameters count and enhance the extraction of deep-layer information. The pyramid module based on Efficient Channel Attention (ECA) is utilized to obtain different receptive fields. And that model can overcome the limitation of the U-Net model with a single receptive field and improve the segmentation capability for targets of different sizes. A deep supervision module with multi-scale output fusion is designed to extract detailed information about liver with high quality. The proposed method is tested on the Liver Tumor Segmentation (LiTS) dataset for liver segmentation, achieving a Dice Similarity Coefficient (DSC) of 92.20% for liver segmentation. Compared to existing models, the proposed method demonstrates higher accuracy in liver segmentation.

**KEYWORDS-** Liver Segmentation, Deep Learning, CT Images, EAS, Liver Tumor Segmentation Dataset

## I. INTRODUCTION

The liver is a vital organ essential for human, and liver cancer has become a frequently occurring type of cancer all around the world. According to the latest data released by the World Health Organization, a significant number of individuals die each year due to liver cancer [1]. In the field of medicine, the primary method for screening liver diseases involves the use of Computed Tomography (CT) scans. In clinical practice, CT scan is an important basis for judging the patient's liver condition and the extent of its development, as well as whether there is a tumor. Therefore, segmenting the patient's liver region from CT images is of great significance to improve doctors' diagnostic efficiency and is very important to save people's life. These CT scans generate multiple consecutive layered images, which are then manually or semi-manually delineated by radiologists to identify the liver or areas of pathology for further diagnosis and treatment planning[2]. However, manual or semi-manual

delineation of the liver or lesion areas is a process taking up a lot of time and is inclined towards subjectivity, and even worse, leading to potential errors. Therefore, automatic segmentation techniques for the liver and tumors is becoming a more and more popular research topic in the current medical field. Because automatic segmentation not only speed up the examination and save doctor's time which they can spend on communicate with patients to gather more information, but also can improve the accuracy by fed with more data and doing more tasks. However, the low contrast between the surrounding organs and tissues and liver in CT results makes it a highly challenging task that the automatic and robust separating the liver out from CT scans. In recent years, deep learning techniques have been broadly used in segmentation medical image tasks, with an increasing number of researchers adopting deep learning technology for liver segmentation tasks[3]. Currently, cutting-edge liver segmentation methods are mostly based on Fully Convolutional Networks (FCN). [4]. The core of FCN is based on an encoder-decoder structure that performs pixel-wise classification of images, allowing it to input images of arbitrary sizes and produce outputs of the same size. Building upon FCN, Ronneberger et al. [5] introduced a fully symmetric Ushaped network known as U-Net. Its main innovation involves fusing each upsampling operation with the corresponding feature extraction segment to capture more low-level semantic information. U-Net achieves relatively accurate results with minimal training data. This network's drawback is that the downsampling operation results in a significant loss of semantic information. To address this issue, Liu et al. [6] designed a network for 3D medical image segmentation called VNet, incorporating the residual connection concept from ResNet[7]. It extracts liver features from the encoder and generates full-resolution output from the decoder. Although this method improves segmentation accuracy, the large parameter count of 3D convolutions requires longer training times. To alleviate this concern, Li et al. [8] proposed H-DenseUNet, blending features from 2D DenseUNet and 3D DenseUNet to expedite the convergence of 3D DenseUNet. Additionally, dense connection blocks are embedded in U-Net to further enhance liver and tumor segmentation accuracy. Furthermore, Han et al. [9]

proposed a method that stacks multiple 2D slices into 3D information for liver and tumor segmentation. This approach combines U-Net's skip connections with ResNet's residual connections and leverages the input of consecutive single slices to provide 3D contextual information. However, the obtained 3D contextual information is limited. In the context of segmenting multi-scale tumor features, Jin et al.[10] introduced RA-UNet, using a 3D U-Net as the base network. It incorporates attention mechanisms in the network's deep layers to achieve liver and tumor segmentation. While this method improves segmentation accuracy to some extent, the deep features become increasingly abstract, limiting the enhancement.

As shown in the above text, training the network solely using the U-Net or ResNet architecture has limitations, including network performance degradation and low utilization of spatial features and extra time costing on training, resulting in lower segmentation accuracy. To address these issues, we employ a pyramid module based on Efficient Channel Attention (ECA) to obtain different receptive fields, overcoming the limitations of a single receptive field in the U-Net model and improving segmentation capabilities for targets of different sizes. Additionally, we design a deep supervision module for multi-scale output fusion, enhancing the segmentation task performance of the U-Net structure.

## II. LIVER SEGMENTATION ALGORITHM INTRODUCTION

With the fast step-forward of deep learning technology, segmentation of medical image based on deep learning is continually improving. Compared with traditional image segmentation algorithms, deep learning algorithms extract features of images to a deeper extent, have better segmentation accuracy, and do not require a lot of manual analysis and parameter adjustment. And segmentation based on deep learning technology is often more universal and able to cover more situations than traditional algorithms. Using deep learning algorithms to replace traditional image segmentation algorithms usually can achieve better segmentation results.

### A. The Network Structure Of U-Net

The U-Net network structure is an improvement based on the Fully Convolutional Network (FCN) and is mainly consist of three parts: the encoder, the decoder, and skip connections. As illustrated in Figure 1, the network is divided into two parts: the left part represents the encoding phase, or the contraction process, while the right part corresponds to the decoding phase, or the

expansion path. The contraction path primarily employs a convolutional network structure. And the path is consisting of two repeated  $3 \times 3$  convolutional kernels. And each convolutional kernel uses the Rectified Linear Unit (ReLU) activation function, and has a  $2 \times 2$  max-pooling layer with a downsampling with step size of 2. At each downsampling stage, as a result of every pooling operation, the image gets a reduction to half of its preceding size, accompanied by a doubled increase in the number of feature channels. For example in Figure 1, the original image size is  $572 \times 572$ , after the first time of max pooling, it become  $284 \times 284$  and similar to the later layers. Until the image complete all the compression to  $32 \times 32$ . In this process, though the size of image is continually be compressed, the features are extracted layer by layer and will be used in the expansion path.

The expansion path is the inverse process of the contraction path, involving deconvolution. In each step, the feature maps undergo upsampling, where each upsampling (Up) operation doubles the size of the image and reduces the corresponding feature channel number to half of its previous value. Finally, a  $1 \times 1$  operation is performed, enabling the network model to analyze and infer pixel-level information. In the encoding phase, U-Net reduces the image size through convolution and downsampling, extracting shallow-level feature information. In the decoding phase, deconvolution and upsampling are employed to capture deeper-level features. Through a skip connection mechanism, the feature maps obtained in the encoding and decoding phases are overlaid, combining both shallow and deep-level features to refine the image. Ultimately, segmentation predictions are made based on the acquired feature maps.

### B. Alterations to the U-Net Structure

In this paper, based on the U-Net architecture, we introduce the use of depth-wise separable convolution modules as a replacement for standard convolution modules. This modification significantly reduces the parameters count in the U-Net Structure. Furthermore, we incorporate an efficient attention mechanism, the Efficient Channel Attention (ECA) module, extracted from the ECA-Net architecture. Combined with a pyramid module, we propose the EAS (ECA-Attention and Separable convolution) feature extraction module, aiming to enhance the extraction of detailed information related to liver target. Finally, we introduce a multi-scale output fusion module to better capture complex features and effectively improve the model's performance.

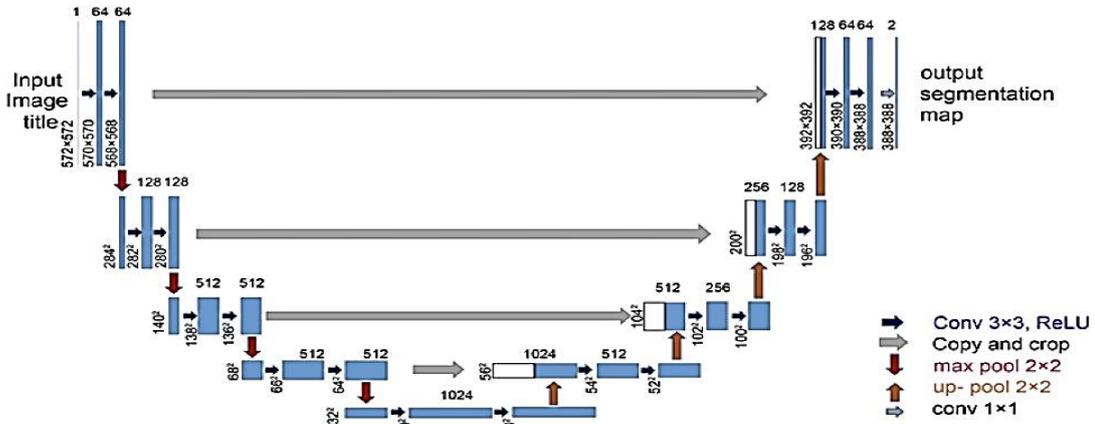


Figure 1: Network structure of U-Net

• **Introducing Depth-wise Separable Convolution Module**

The depth-wise separable convolution module [11] is a specialized convolution operation. There are two steps in the convolution process, which are depth-wise separable convolution and point-wise convolution. In the first step, depth-wise separable convolution, each channel of the input image are convolved separately, rather than convolving all channels together. This approach reduces the number of parameters, which contributes to boosting the model’s computational efficiency. The second step is point-wise convolution convolving the output of depth-wise convolution to extract further features. As shown in Figure 2, the conventional convolution module is similar, while the depth-wise separable convolution module, as depicted in Figure 3, includes Batch Normalization (BN) and an ReLU activation function layer after the depth-

wise and point-wise layers. However, using ReLU as the activation function has the drawback. Because it has a characteristics that setting negative values to 0, which will result in losing some feature information. To improve this, LeakyReLU [12] is employed as a substitute for ReLU. LeakyReLU closely resembles ReLU; however, unlike ReLU, it outputs a small negative value for negative inputs instead of zero. This design mitigates the “dying ReLU” issue, preventing certain neurons from remaining inactive throughout the training process. This property becomes particularly relevant after point-wise convolution. The h-swish function is used instead of ReLU as the activation function to reduce the issue of vanishing gradients caused by an increase in the number of network layers, as illustrated in Figure 4. The advantages of this module include improved computational

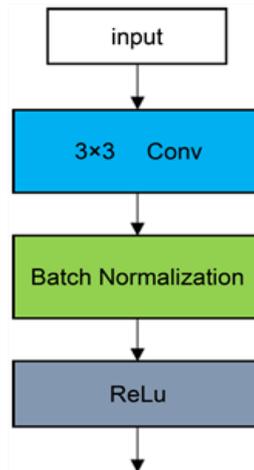


Figure 2: Standard Convolutional Module

efficiency, reduced parameter count, enhanced extraction of deep-level information, and increased model accuracy. The h-swish function is a simplified variant of the swish activation function, and the computation of swish is represented by Equation:

$$Swish(x) = x \cdot sigmoid(\beta \cdot x)$$

$$sigmoid = \frac{1}{1 + e^{-x}}$$

$B$  is a constant or trainable parameter. Swish possesses characteristics of being unbounded above and bounded below, smooth, and non-monotonic. In comparison to the ReLU activation function and its derivatives, Swish activation function significantly improves the accuracy of target detection in convolutional neural networks. However, swish function faces challenges in terms of computational complexity, and incorporation of the h-swish function efficiently tackles the computational constraints associated with the swish function. The

computation of the h-swish function is represented by Equation:

$$h - Swish[x] = x \cdot \frac{ReLU6(x + 3)}{6}$$

$$ReLU6[x] = \min(\max(0, x), 6)$$

The adjustable parameter in swish enables fine-tuning of the activation function to enhance information flow, promoting more gradual gradients. This, in turn, facilitates smoother optimization of the landscape, leading to improved generalization and faster convergence.

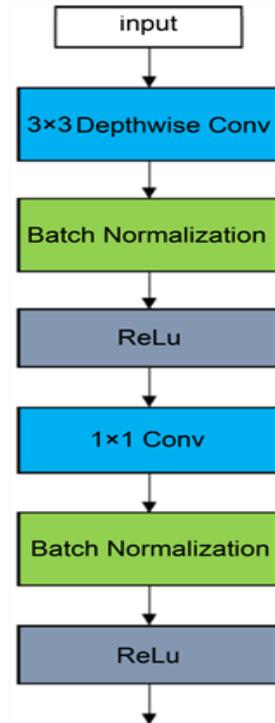


Figure 3: Deep Separable Convolutional Module

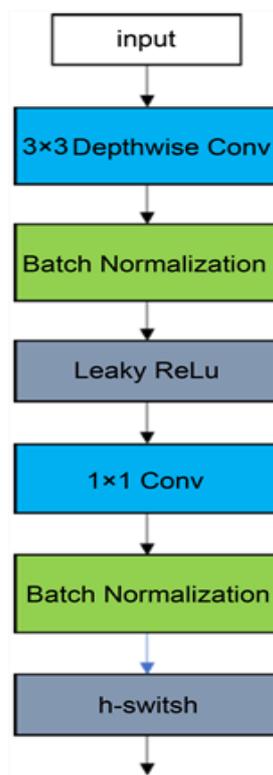


Figure 4: Optimized Deep Separable Convolutional Module

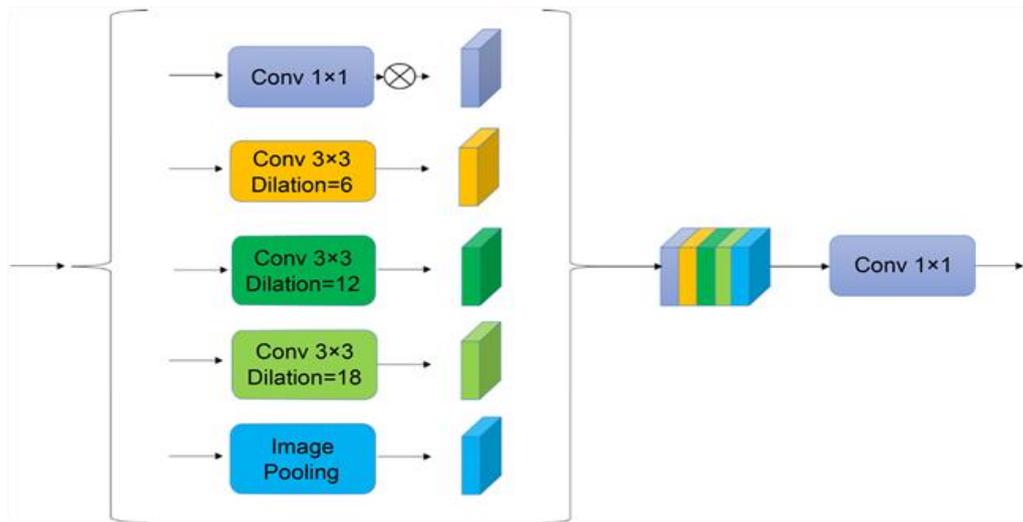


Figure 5: Pyramid Modules Based On Attention Mechanism

#### • *The Pyramid Module Based on Attention Mechanism*

The Dilated Spatial Pyramid Pooling module proposed by Chen et al. [13] in the DeepLabV3 algorithm. The ASPP (Atrous Spatial Pyramid Pooling) module consists of a set of parallel atrous convolutions with multiple dilation rates, enhancing the recognition of multi-scale targets by adjusting the convolutional kernel's receptive field. ECA-Net [14] is an efficient channel attention mechanism that improves the original SE-Net module with minimal parameters while enhancing the network's detection performance. The ASPP module not only introduces a global average pooling path to obtain image-level features but also employs multiple atrous convolution paths, minimizing the loss of important edge details in the target due to downsampling to some extent. Importantly, it can capture features at multiple scales in the image. However, the first path in the parallel structure of the ASPP module does not utilize atrous convolutions to obtain features in multi-scale. Instead, it adjusts the channel number of the input feature map using a  $1 \times 1$  convolutional kernel, which cannot effectively filter the features. The previously mentioned ECA-Net network model belongs to the channel attention mechanism, adjusting the importance of features by applying attention to feature channels. In this paper, we integrate it with the ASPP module to propose the EAS (ECA-ASPP) module. We embed the EAS module into the encoding side of the U-Net network, enhancing the network's ability to extract soft bone features at multiple scales. As shown in Figure 5, the first convolution is  $1 \times 1$  and no dilation, then the second convolution is  $3 \times 3$  and dilation is 6 and so on.

#### • *Multi-scale Output Fusion Module*

The concept of multi-scale output fusion involves integrating information from different scales to enhance the overall understanding and representation of features in a neural network. This module aims to combine outputs from multiple layers or pathways, each capturing information at a distinct scale. The fusion process ensures that the model has the ability to effectively leverage both fine-grained level details and coarse-level features,

leading to improved performance in tasks such as object recognition, segmentation, or classification.

For U-Net networks adopting an encoder-decoder structure, only the final stage outputs the ultimate model's predicted segmentation map. That map of the final model is obtained by fusing the the last skip connection's output feature maps. Although the last stage combines low-level features with high-level features that have undergone fusion and compression, possessing good spatial and pixel category information, the encoding side of the model undergoes multiple operations like convolution and pooling, reducing the model's parameters while losing some feature information. Additionally, in the decoding side, multiple upsampling operations lead to the restoration of the feature map size but also result in information loss, both of which contribute to a decrease in prediction accuracy.

Hence, a multi-scale output fusion module [15] is proposed, as illustrated in Figure 6. It can merge outputs from different scales to enhance the model's accuracy and precision. This module assists the model in capturing complex features more effectively and improves overall performance. Moreover, the multi-scale output fusion module aids the model in better handling inputs of different scales, thereby improving accuracy and precision. In terms of feature fusion methods, there are generally two approaches: additive fusion and concatenative fusion. Additive fusion involves adding corresponding elements at each position of the feature maps from the upper and lower pathways. However, this method may lead to the loss of original feature information and may not reflect the complementary nature of features. Concatenative fusion, on the other hand, directly concatenates the feature maps from the upper and lower pathways, avoiding the information loss caused by additive fusion. It merges different channels, synthesizing features obtained earlier to produce subsequent features.

#### *C. Loss Function Choice*

During network training, the choice of the loss function is a crucial aspect. The smaller the value of the loss function, the closer the predicted values are to the ground truth. Cross Entropy (CE) is the most widely used loss

function in medical image segmentation. However, in cases of segmentation class imbalance, smaller targets are often overlooked.

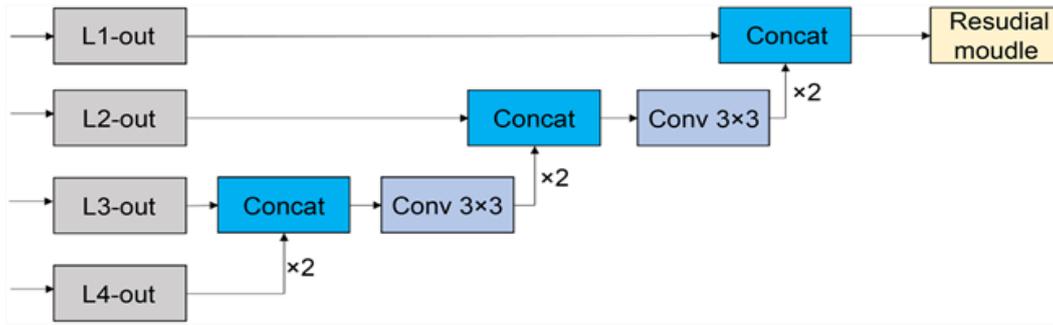


Figure 6: Deep supervision module of multi-scale output fusion

As evident from the figure, this paper deals with small target segmentation, where cartilage occupies a smaller proportion of the image. Dice Loss (DL) is a loss function derived from the Dice coefficient, calculated based on the difference in overlap (F) between the predicted and ground truth results, where Pre represents the predicted image and GT represents the ground truth image. Therefore, this paper combines both, and the final loss function consists of two parts.



$$Diceloss = 1 - \frac{2|Pre \cap Gt|}{|Pre| + |GT|}$$

$$Loss = CEloss + Diceloss$$

### III. RESULTS AND ANALYSIS OF EXPERIMENTS

#### A. Environment of Experiments

The experimental environment is based on operating system the Ubuntu , programming language Python 3.10, and deep learning framework PyTorch 2.0.0 .The hardware configuration includes an Intel Core i9 14900HX CPU, 64GB of RAM, and an RTX 4090 graphics card, with CUDA 11.8 for accelerated computing. The optimizer uses Stochastic Gradient Descent (SGD) and the learning rate is set to 0.01. And the optimizer has a weight decay of 0.0001 and a momentum coefficient of 0.9. The batch size is set to 32, and the training epochs predefined number is 50.

#### B. Data Set and Preprocessing

The dataset was obtained from the Liver Tumor Segmentation (LiTS) competition initiated by the Medical Image Computing and Computer Assisted Intervention Society (MICCAI) in 2017 [16]. LiTS comprises 200 contrast-enhanced 3D abdominal CT scan images from six different hospitals. For supervised training, MICCAI provided manually annotated masks drawn by experienced radiologists for liver and tumor regions. The dataset differs significantly in spatial resolution, image quality, and visual aspects. The in-plane resolution is

fixed at  $512 \times 512$  pixels, with pixel distances ranging from  $0.6 \text{ mm} \times 0.6 \text{ mm}$  to  $1.0 \text{ mm} \times 1.0 \text{ mm}$ . The slice spacing ranges from  $0.45 \text{ mm}$  to  $6.0 \text{ mm}$ , and the number of slices for different cases varies from 42 to 1026. We sliced the nii files into 2D images along the z-axis. Simultaneously, slices containing non-liver regions or with liver area less than 1.5% were not retained during the segmentation process.

In contrast to natural images, CT scan images store pixel values in Hounsfield units (Hu), with pixel values ranging between  $[-3000, 3000]$ . In this experiment, the suitable intensity values within the range of  $[-200, 250]$  were obtained by analyzing the window width and window level values of CT images and the liver region. Pixels greater than 250 Hu were set to 250, and pixels less than -200 Hu were set to -200, allowing for a clearer observation of the liver region to be segmented. After adjusting the window width and window level, the liver region becomes more clearly visible.

Regarding dataset partitioning, this paper randomly divides the 130 cases into a training set and a validation set in an 8:2 ratio, with the remaining 70 cases forming the test set. Additionally, to prevent model overfitting, the paper employs random cropping to augment the data. The cropped size is set to  $32 \times 256 \times 256$ . Additionally, data augmentation was performed by flipping the images vertically and horizontally, as well as applying random rotations at specific angles to enhance the dataset.

#### C. Evaluation Metrics

To accurately measure the segmentation performance of the network on liver CT medical images, this paper adopts Dice Similarity Coefficient (DSC), precision, recall and Hausdorff distance as evaluation metrics.

##### • Dice Similarity Coefficient (DSC)

The Dice Similarity Coefficient (DSC) stands out as the widely employed metric in segmentation tasks for assessing the resemblance between two samples, with a range of  $[0, 1]$ , as shown in formula, where Pre represents the network's predicted image, GT represents the ground truth image. The bigger of the DSC is, the better performance of the segmentation is.

$$DSC = \frac{2|Pre \cap GT|}{|Pre| + |GT|}$$

- **VOE (Volumetric Overlap Error)**

VOE (volumetric overlap error) represents the error rate, measured in percentage where ( $V_{false}$ ) represents the volume where the predicted and true results do not overlap, and ( $V_{true}$ ) represents the total true volume. 0% indicates complete segmentation, while 100% indicates no overlap between the predicted and ground truth results.

Its formula is:

$$VOE = 1 - \frac{2 \times (V_{true} \cap V_{pred})}{V_{true} + V_{pred}}$$

- **ASD(Average Symmetric Surface Distance)**

ASD (average symmetric surface distance) represents the average surface distance between the predicted and ground truth results in symmetric positions, measured in millimeters, where 0 mm indicates perfect segmentation, where ( $N$ ) represents the number of points, ( $d_i$ ) represents the surface distance from the ground truth, and ( $d'_i$ ) represents the surface distance from the prediction.. Its formula is:

$$ASD = \frac{1}{N} \sum_{i=1}^N |d_i - d'_i|$$

#### D. Experiments and Comparison

This study addresses the issue of unsatisfactory performance of the U-Net model in the segmentation of liver CT images. The network model was improved, leading to successful model training. Figure 7 shows the loss curve of the model on the training set. As seen from Figure 7, when the number of iterations reaches around 47, the loss function value essentially stabilizes, settling around 0.0273. This indicates that the model has essentially converged. After approximately 47 iterations, the loss function value stabilizes around 0.0273, representing the completion of training. Also Yang introduce a way to visualize of time series data, [17]

Figure 9 is the result obtained after segmentation of the test set Figure 8 by the trained model.

To further verify whether the improved network enhances segmentation accuracy, calculations were performed for the following metrics. As shown in Table 1, compared to other network models, the proposed improved network model in this study exhibits certain advantages in various metrics on the same test set. This indicates that the enhanced model in this study achieves satisfactory segmentation results for liver.

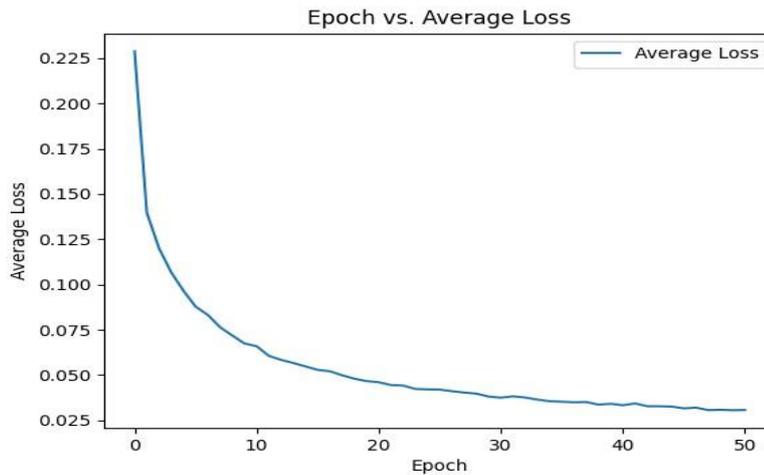


Figure 7: Training set loss curve

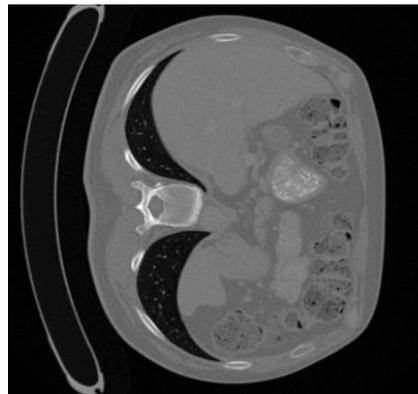


Figure 8: Original Image

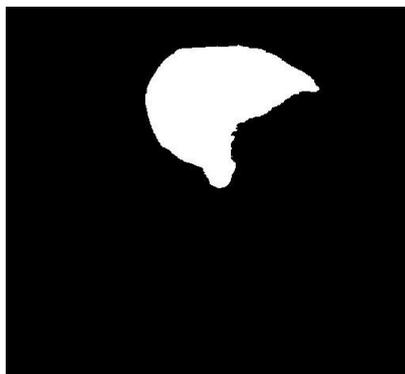


Figure 9: Our Segmentation Result

Table 1: Comparison of different components in the liver segmentation

Model	Liver Segmentation DSC(%)	VOE(%)	ASD(mm)
U-Net 2D	89.13	90.54	8.13
VNet 3D	89.40	90.68	7.53
Ours	92.20	93.05	7.65

#### IV. CONCLUSION

The paper proposes an end-to-end EAS U-Net based on the U-Net network. By replacing standard convolutional modules with depth-wise separable convolutional modules, the number of parameters is reduced, leading to improved computational efficiency. The model utilizes an ASPP (Atrous Spatial Pyramid Pooling) based on ECA (Efficient Channel Attention) to mitigate the loss of target edge information caused by downsampling operations. Additionally, a multi-scale output fusion module is designed to complement feature information. Through comparative experiments with different network models, the proposed model in this paper achieves better segmentation accuracy in liver segmentation.

#### CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

#### REFERENCES

1. J Ferlay, HR Shin, F Bray, D Forman, C Mathers, and DM Parkin. Estimates of worldwide burden of cancer in 2008: Globocan 2008. *International Journal of Cancer*, 127:2893–2917, 2010.
2. Müller S A, Maier-Hein L, Mehrabi A, Pianka F, Rietdorf U, Wolf I, Grenacher L, Richter G, Gutt C N, Schmidt J, Meinzer H P, and Schmied B M. Creation and establishment of a respiratory liver motion simulator for liver interventions. *Medical Physics*, 34:4605–4608, 2007.
3. Dou Qi, Chen Hao, Jin Yueming, Yu Lequan, Qin Jing, and Pheng-Ann Heng. 3d deeply supervised network for automatic liver segmentation from ct volumes. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Athens, 17-21 October 2016:149–157, 2016.
4. J Long, E Shelhamer, and T Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:640–651, 2015.
5. Ronneberger Olaf, Fischer Philipp, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention— MICCAI 2015*, Springer, Cham:234–241, 2015.
6. F. Milletari, N. Navab, and S.A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571, Stanford, October 2016.
7. K.M. He, X.Y. Zhang, S.Q. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, Las Vegas, June 2016.
8. X. Li, C. Hao, X. Qi, D. Qi, C.W. Fu, and H. Pheng-Ann. H-denseunet: Hybrid densely connected unet for liver and liver tumor segmentation from ct volumes. *IEEE Transactions on Medical Imaging*, 37:2663–2674, 2017.
9. X. Han. Automatic liver lesion segmentation using a deep convolutional neural network method. *ArXiv preprint*, 2017. arXiv:1704.07239.
10. Q. Jin, Z. Meng, C. Sun, L. Wei, and R. Su. Ra-unet: A hybrid deep attentionaware network to extract liver and tumor in ct scans, 2018.
11. Howard Andrew G, Zhu Menglong, Chen Bo, and et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017:1382–1391, 2017.
12. Maas Andrew L, Hannun Awni Y, and Ng Andrew Y. Rectifier nonlinearities improve neural network acoustic models. *Proceedings of the 30th International Conference on Machine Learning*, 2013.
13. Chen Liang-Chieh, Zhu Yukun, Papandreou George, Schroff Florian, and Adam Hartwig. Encoder-decoder with atrous separable convolution for semantic image segmentation. *European Conference on Computer Vision (ECCV)*, Springer, Cham:833–851, 2018.
14. Wang Qilong, Wu Banggu, Zhu Pengfei, Li Peihua, Zuo Wangmeng, and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. *CVPR 2020*, 2019.
15. Wei LU, Dan LIU, Min SHAO, and Yangdong WU. Improved mask r-cnn multitarget detection and

- segmentation for autonomous driving in complex scenes. Journal of Computer Engineering & Applications, 2021, Vol 57, Issue 24:p234, 2023.
16. Bilic Patrick, Christ Patrick, Bran Li Hongwei, Eugene Vorontsov, and et al. The liver tumor segmentation benchmark (lits). Medical Image Analysis (2022), page Pg. 102680, 2019.
  17. H YANG, T LI, and X CHEN. Visualization of time series data based on spiral graph. Journal of Computer Applications, 2017.